Exhibit 75

Annals of Internal Medicine IDEAS AND OPINIONS

Responding to Reviewers and Editors About Statistical Significance Testing

David A. Savitz, PhD; Lauren A. Wise, MSc, ScD; Julia C. Bond, MPH; Elizabeth E. Hatch, MS, PhD; Collette N. Ncube, DrPH, MPH, MS; Amelia K. Wesselink, PhD; Mary D. Willis, PhD, MPH; Jennifer J. Yland, PhD; and Kenneth J. Rothman, DrPH

The deficiencies of statistical significance testing as a framework for presenting and interpreting research findings have been emphasized for decades (1–3). These deficiencies include the dichotomization of results that are on a continuum; inaccurate interpretation of results that are not statistically significant as supporting the null hypothesis; incentive to alter the analysis to attain statistical significance; and loss of important details about the magnitude, pattern, and precision of observed associations. A complete presentation of research findings requires information on both the magnitude and precision of effect estimates along with evidence about potential biases. Statistical significance testing conflates precision with effect size rather than considering each separately. Dichotomizing findings into "significant" and "not significant" reduces information, obscures biases, and often leads to misinterpreting strong associations as null and nearly identical results as conflicting (4).

The statistics community (5) and medical journal editors (6) have long recognized this problem and have begun to more forcefully emphasize quantitative interpretation of results rather than assignment into discrete bins. A correctly interpreted CI provides information about both the magnitude and precision of effects, indicating a range of possible parameter values that are statistically compatible with the data (7).

For example, in a recently submitted manuscript, we noted that <*x exposure*> was associated with a hazard ratio of 1.30 with a 95% CI of 0.96 to 1.75 for <*y outcome*>. Using dichotomous significance testing for inference could lead to an interpretation of "no effect." A more appropriate interpretation would be that <*x exposure*> was modestly associated with <*y outcome*>, but the data were statistically consistent with parameter values ranging from little or no effect to a considerable increase in risk. The boundaries of the CI should not be viewed as sharp demarcations between plausible and implausible but should rather suggest the range of values most consistent with the data, with the degree of consistency greatest for values in the center of the interval and falling off steadily and continuously for values farther away.

Reviewers or editors sometimes contradict official recommendations against dichotomous statistical significance testing. Authors may be asked to provide statistical significance tests when CIs have been presented and to refer to associations that are not statistically significant as though they were null findings. Here, we suggest 2 approaches to how an author trying to avoid dichotomous statistical significance testing can address editorial requests to use it.

The first approach is to clarify the philosophy—that is, indicate clearly in the methods section that significance tests are not being used to interpret results. For example, authors could state, "Our approach to interpreting data is based on an evaluation of the magnitude, direction, and precision of the effect estimates rather than binary significance testing." Explain why a more nuanced approach to the interpretation of data is preferable to interpreting results on the basis of dichotomization using tests of statistical significance. This explanation can draw on publications (8, 9) that emphasize that evidence falls on a continuum and any dichotomization degrades the information in the data.

A second approach is to cite recent, authoritative sources explaining that quantitative interpretation is consistent with recommendations from the American Statistical Association (5), the International Committee of Medical Journal Editors (6), and current textbooks on epidemiologic methods (10). A recent commentary by Amrhein and colleagues (3) provides a clear explanation that is applicable across disciplines, with more than 800 scientists indicating their agreement.

Consider the following example in which we examined the association of an exposure with 2 distinct outcomes (disease A and disease B) and submitted the resulting manuscript for peer review. The submitted manuscript summarized the results in a **Table** and stated, "An elevated risk of [disease B] was found, but with limited precision." The reviewer suggested removing any mention of imprecision and recommended concluding that "there was no significant increase in the risk of [disease B]."

In the original manuscript, the authors attempted to incorporate considerations of the magnitude, precision, and pattern of the association. The reviewer was asking for a statement such as "exposure was significantly related to disease A but not disease B." This approach would obscure that both disease A and B show dose–response gradients, with a *P* value for linear trend of 0.0015 for disease A and 0.0853 for disease B and a somewhat steeper gradient for disease B.

In our response, we argued against making a definitive declaration about either disease A or B because decisions about causal effects and clinical recommendations should not be made solely on the basis of this study's results. Furthermore, there is no reason to make a sharp distinction between the 2 effects because there

This article was published at Annals.org on 20 February 2024.

Table. Results From a Submitted Manuscript Showing an Approach to Interpretation

Variable	Cases, *n*	Absolute Risk, %	Odds Ratio	
			Unadjusted	Adjusted (95% CI)
Disease A				
Overall	4237	0.92	–	–
No exposure	675	1.01	1.00	1.00
Low exposure	1330	1.02	1.00	1.16 (1.05–1.27)
Medium exposure	1248	0.95	0.94	1.18 (1.07–1.30)
High exposure	984	0.75	0.74	1.22 (1.08–1.38)
Disease B				
Overall	328	0.07	–	–
No exposure	48	0.07	1.00	1.00
Low exposure	106	0.08	1.12	1.28 (0.91–1.80)
Medium exposure	103	0.08	1.09	1.40 (0.97–2.00)
High exposure	71	0.05	0.75	1.44 (0.93–2.25)

is separate interest in each outcome. An informed assessment must consider all findings along with the quality of exposure and disease ascertainment, susceptibility to a range of biases (for example, selection and information biases), and effectiveness of control for confounding. To conclude that disease A was significantly associated with exposure and disease B was not would leave the reader with an incomplete and misleading synopsis of the findings.

The publication of a manuscript is generally perceived as a product of the authors as well as a reflection of the journal's quality. Guidance from editors and reviewers generally improves the manuscript and should always be given serious consideration. Nevertheless, well-reasoned resistance is appropriate when reviewers or journal editors request methodological revisions contrary to expert recommendations. The path of least resistance would be to concede to journal stipulations, particularly those that come from editors rather than reviewers. However, such concessions may do a disservice to science, harm authors, and impede progress for others.

From Department of Epidemiology, Brown University School of Public Health, Providence, Rhode Island (D.A.S.); and Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts (L.A.W., J.C.B., E.E.H., C.N.N., A.K.W., M.D.W., J.J.Y., K.J.R.).

Financial Support: None.

Disclosures: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M23-2430.

Corresponding Author: David A. Savitz, PhD, Department of Epidemiology, Brown University School of Public Health, 121 South Main Street, Box G-S121, Providence, RI 02912; e-mail, david_savitz@brown.edu.

Author contributions are available at Annals.org.

Ann Intern Med. doi:10.7326/M23-2430

References

1. **Rothman KJ.** A show of confidence. N Engl J Med. 1978;299:1362-1363. [PMID: 362205] doi:10.1056/NEJM197812142992410
2. **Greenland S, Senn SJ, Rothman KJ, et al.** Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31:337-350. [PMID: 27209009] doi:10.1007/s10654-016-0149-3
3. **Amrhein V, Greenland S, McShane B.** Scientists rise up against statistical significance. Nature. 2019;567:305-307. [PMID: 30894741] doi:10.1038/d41586-019-00857-9
4. **Lash TL.** The harm done to reproducibility by the culture of null hypothesis significance testing. Am J Epidemiol. 2017;186:627-635. [PMID: 28938715] doi:10.1093/aje/kwx261
5. **Wasserstein RL, Lazar NA.** The ASA statement on *p*-values: context, process, and purpose. Am Stat. 2016;70:129-133. doi:10.1080/00031305.2016.1154108
6. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. Updated January 2024. Accessed at www.icmje.org/icmje-recommendations.pdf on 23 January 2024.
7. **Greenland S.** Connecting simple and precise *P*-values to complex and ambiguous realities (includes rejoinder to comments on "Divergence vs. decision *P*-values"). Scandinavian Journal of Statistics. 2023;59:899-914. doi:10.1111/sjos.12645
8. **Farland LV, Correia KF, Wise LA, et al.** *P*-values and reproductive health: what can clinical researchers learn from the American Statistical Association? Hum Reprod. 2016;31:2406-2410. [PMID: 27664212] doi:10.1093/humrep/dew192
9. Rothman KJ. Random error and the role of statistics. In: Epidemiology: An Introduction. 2nd ed. Oxford Univ Press; 2012:148-163.
10. Lash TL, VanderWeele TJ, Haneuse S, et al. Modern Epidemiology. 4th ed. Lippincott Williams & Wilkins; 2021.

Author Contributions: Conception and design: D.A. Savitz, L.A. Wise, A.K. Wesselink, J.J. Yland, K.J. Rothman.
Drafting of the article: D.A. Savitz, L.A. Wise, J.C. Bond, A.K. Wesselink, J.J. Yland, K.J. Rothman.
Critical revision of the article for important intellectual content: D.A. Savitz, L.A. Wise, J.C. Bond, E.E. Hatch, C.N. Ncube, A.K. Wesselink, M.D. Willis, J.J. Yland, K.J. Rothman.
Final approval of the article: D.A. Savitz, L.A. Wise, J.C. Bond, E. E. Hatch, C.N. Ncube, A.K. Wesselink, M.D. Willis, J.J. Yland, K. J. Rothman.
Statistical expertise: A.K. Wesselink, K.J. Rothman.